*THE AMINO ACID SEQUENCE OF BOVINE CARBOXYPEPTIDASE A**

By Ralph A. Bradshaw,† Lowell H. Ericsson, Kenneth A. Walsh, and Hans Neurath

DEPARTMENT OF BIOCHEMISTRY, UNIVERSITY OF WASHINGTON, SEATTLE

Communicated June 11, 1969

Abstract.—The amino acid sequence of the four fragments produced by treatment of bovine carboxypeptidase A with cyanogen bromide has been completed. The alignment of these fragments, previously established by peptic digest of the whole protein, allows for the description of the complete primary structure of the molecule. A comparison of the proposed functional residues, identified by X-ray diffraction analyses, has either confirmed their assignments or provided their correct identity. The functional and structural residues of principal importance include Arg 145, Tyr 248, and Glu 270 as the binding site of the substrate carboxyl group, the proton donor, and the nucleophilic moiety, respectively, which were correctly assigned; His 196 as the third zinc ligand and Tyr 265 as the binding site of the α-carboxyl group have been corrected from their original X-ray assignments. The other two zinc ligands, His 69 and Glu 72, were identified previously from chemical and X-ray studies. The assignment of the two half-cystinyl residues and the postulation of the existence of a disulfide bond have been confirmed.

The complete elucidation of the mechanism of action of any enzyme depends ultimately upon the detailed knowledge of the primary and three-dimensional structure of the protein molecule. These features must in turn be translated into a working hypothesis which is consistent with the pertinent chemical and physical properties describing the enzymatic function. As a step toward meeting this goal, the primary structure of bovine carboxypeptidase *A* has now been established. This enzyme has already served as the subject of many investigations, which have recently been reviewed,[1, 2] and to the extent that it is possible, this communication is intended to correlate the features of the amino acid sequence which are most directly related to studies of the function[3] and three-dimensional structure.[4]

The principal route for the solution of the amino acid sequence was through the isolation and characterization of the fragments generated by the action of cyanogen bromide on the protein. This reaction yielded the four fragments expected from the three methionyl residues present in the protein.[5] Isolation of the amino[6] and carboxyl[7] terminal fragments, designated F_N and F_C, respectively, served as the first primary structural information on carboxypeptidase A. In addition, it was possible, through analyses of the F_N fragments isolated from carboxypeptidases A prepared by different methods,[8, 9] to establish the relationship of the α, β, and γ forms of the enzyme. Subsequent purification of these forms by chromatographic techniques[10] has confirmed these assignments. The α form of the enzyme, which has the longest polypeptide chain, possesses an

F_N fragment of 22 residues, while the β and γ forms contain F_N fragments of 20 and 15 residues, respectively.[6, 10]

The F_C fragment contains only six amino acid residues, but is distinguished by the fact that it contains an allotypic replacement site. This variation has been subsequently shown[11] to be linked to two additional sites in the molecule, creating the situation that bovine carboxypeptidase A can occur in two distinct forms which differ at three positions in the sequence. These types, which have been designated Val and Leu forms on the basis of the residues first found to occupy the replacement site in the F_C fragment, are equally distributed between the activation forms described above. As a result of the heterogeneity due to activation conditions and that due to allelomorphism, bovine carboxypeptidase A has been shown to exist in six unique forms, i.e., A_α^{Val}, A_α^{Leu}, A_β^{Val}, A_β^{Leu}, A_γ^{Val}, and A_γ^{Leu}.

The isolation and characterization of the remaining two fragments, designated F_I and F_{III} from chromatographic experiments, was accomplished by gel filtration.[12] The summation of the amino acid compositions of the four fragments F_N, F_C, F_I, and F_{III} to give the composition of the native enzyme and their alignment in the order F_N–F_{III}–F_I–F_C established that the α form of the enzyme contains 307 residues with a molecular weight of 35,268.

The structure of the F_{III} fragment, reported previously,[1, 2] was deduced from the tryptic, chymotryptic, and thermolytic peptides.[13, 14] The fragment contains 81 residues and, in combination with the F_N fragment, supplied the sequence of the first 103 residues. A comparison[1, 4] of this structure with that obtained from X-ray diffraction studies, after suitable correction, indicated only minor disagreements at four sites.

The completion of the structure of the F_I fragment[15] supplied the last piece of information necessary to give the entire primary structure of the molecule. The larger size and the greater difficulty of obtaining soluble peptides from this fragment necessitated the use of five different enzymic digests before the requisite alignments and overlaps were obtained. In all, peptides from digestions by trypsin, chymotrypsin, thermolysin, pepsin, and Nagarse were used to build the final structure. The data from these experiments, and those described above, have been assembled to yield the structure shown in Figure 1.

In addition to the structures of F_N, F_C, and F_{III}, several segments of the F_I fragment have been described previously. In each case, complete agreement was found between these segments and the final structure. These regions included residues 138–144 and 152–165 containing the half-cystinyl residues,[16] residues 247–250 containing the site of iodination reported by Roholt and Pressman,[17] and residues 280–293, the segment homologous to the thiol peptide of carboxypeptidase B.[18] In view of these residues, as well as several other segments that have been communicated privately to W. N. Lipscomb,[4] no attempt has been made to correlate the degree of accuracy of the X-ray analysis in correctly identifying the amino acid side chains in this region of the molecule.

The identification of several amino acid residues in terms of structural and functional properties of the enzyme has been reported by Lipscomb *et al.*[4] In view of the success of these workers in identifying the α-carbon positions of the

FIG. 1.—Amino acid sequence of bovine carboxypeptidase Aα. The vertical arrows refer to the three sites of activation which yield the α, β, and γ forms of the enzyme.

main chain as well as the nature of many of the residue side chains, it appears certain that most of these residues have been correctly assigned in terms of their position in the chain. The establishment of the complete sequence by chemical means now affords the opportunity for unambiguously identifying the character of all of the side chains, thus solidifying the previously deduced conclusions regarding the role of these residues in the function of the enzyme.[3, 4]

The residues identified by Lipscomb *et al.*[4] are summarized in Table 1 in the order that they occur in the sequence of carboxypeptidase A_α. The first three residues listed occur in the F_{III} portion of the molecule and have been rigorously identified by both chemical and X-ray analyses.[1, 4] Residue 127, identified by X-ray analyses as Glx has been established to be Arg, suggesting that the possible alternative mechanism involving this residue is unlikely. Residues 138 and 161 were identified by X-ray analysis as half-cystinyl residues joined in space by a disulfide linkage. The assignment of these residues has proved correct, and postulation of the disulfide bond has been consequently substantiated by chemical means.[19] It may, therefore, be concluded that the two half-cystinyl residues in bovine carboxypeptidse A do not participate in the binding of the catalytically essential metal ion.

Residue 145 has been identified as an arginyl residue, and this assignment has been confirmed by the sequence analysis. The postulated role of this residue as the site of interaction of the free α-carboxyl of the substrate with the enzyme remains clearly plausible. Residue 196, which has been indicated as the third zinc ligand, has been established as a histidyl residue. This assignment is in contrast to the Lys (or Glx) identification of the X-ray studies but is consistent with the presence of two nitrogen and one oxygen ligand as argued for by Lipscomb *et al.*[4] from instability constants of model compounds.[20] Thus, the chelation structure which binds the Zn ion of carboxypeptidase A is composed of two his-

TABLE 1. *Identification of possible functional residues in bovine carboxypeptidase A.*

Position in sequence	Identity		Possible function from X-ray analysis*
	X-ray*	Chemical	
69	His	His	Zn ligand
71	Arg	Arg	Substrate binding site
72	Glu	Glu	Zn ligand
127	Glx	Arg	Possible alternate catalytic residue
138	Cys	Cys	Half of disulfide bond
145	Arg	Arg	Substrate binding site
161	Cys	Cys	Half of disulfide bond
196	Lys (or Glx)	His	Zn ligand
198	Tyr	Tyr	Substrate binding site
247	Ile	Ile	Substrate binding site (steric hindrance to some substrates)
248	Tyr	Tyr	Proton donor
249	Glx	Gln	Stabilization of Tyr 248
265	Arg	Tyr	Half of salt linkage to α-carboxyl group
270	Glu	Glu	Nucleophile
279	His (or Phe)	Phe	Substrate binding site; possible alternate catalytic residue
307	Asn	Asn	Carboxyl terminal residue

* Taken from Lipscomb *et al.*[4]

tidyl and one glutamyl residue, based on the evidence contributed by X-ray and chemical sequence analyses.

Residue 198 has been assigned to a tyrosyl residue and attributed to forming part of the substrate binding site, as has residue 279. The assignment of the former has been confirmed, and the ambiguity associated with residue 279 (histidine or phenylalanine) has been resolved. The identification of this residue as phenylalanine, however, removes it from consideration as being involved in a possible alternate mechanism. The group of three residues at positions 247, 248, and 249 has been confirmed in its identification from the X-ray analyses[4] and is in accord with the structure of the peptide isolated by Roholt and Pressman.[17] Additional evidence[21] has also been obtained that tyrosyl residue 248 provides the major site of nitration (although quantitative data are still lacking) and supports the postulated role of this residue as a proton donor in carboxypeptidase-catalyzed hydrolysis of peptide substrates.[3, 4]

Residue 265 has been postulated as the site of internal bonding of the α-carboxyl group of the polypeptide chain in a salt linkage to an arginyl residue. The identification of this residue as tyrosine negates this proposal but in turn allows for the possibility that the interaction with the α-carboxyl group is through a tyrosyl-carboxyl hydrogen bond rather than a salt linkage.

The role of glutamic acid 270 as the principal nucleophilic moiety remains a plausible hypothesis since the sequence analysis confirms the identity of this residue. The nature of the carboxyl terminal residue has already been established.[7]

Thus, the rigorous identification of the proposed functional residues of carboxypeptidase has confirmed, for the most part, the identifications made from the X-ray studies and removed some of the areas of ambiguity. It has correctly established the probable chelate structure which binds the metal ion, has clarified the correct nature of the interaction of the α-carboxyl group, and has eliminated some of the possible alternate mechanisms proposed from the X-ray analysis.

Further examinations of the primary structure have indicated vestiges of gene duplication events accompanying the evolution of this enzyme and have been documented elsewhere.[22]

The completion of the primary sequence of bovine carboxypeptidase A prepares the way for many additional studies. Besides allowing for the final refinement of the three-dimensional structure, it will make possible the extensive chemical and physical studies that are necessary to test the validity of the various postulated mechanisms and the relation of the crystal structure to the solution conformation of this enzyme.

The authors are indebted to Mrs. Brita Moody, Mrs. Milagros Santarin, Mrs. Lynne Maston, Mrs. Merry Jarvis, Mrs. Rosa Mae MacDonald, Mrs. Carol Roesler, Mr. C. V. Shih, Mr. Richard R. Granberg, Mr. Thomas A. Oas, and Mr. John Richardson, without whose able technical assistance this analysis would not have been possible. They should also like to thank their numerous associates, including Drs. Donald R. Babin, Masao Nomoto, N. G. Srinivasin, and Moira Barton, whose contributions to this work were instrumental in its completion. It should be further noted that the exchange of early results with Prof. William N. Lipscomb and his co-workers served as a useful guide in

many of the initial aspects of this work and is duly acknowledged. The extensive exchange of information with Prof. Bert L. Vallee and his co-workers has been of substantial aid throughout this work.

* This investigation has been supported by grants from the National Institutes of Health (GM 04617 and GM 15731-01), the National Science Foundation (GB 4990), and the American Cancer Society (P-79K).

† Present address: Department of Biological Chemistry, Washington University School of Medicine, St. Louis, Missouri, 63110.

[1] Neurath, H., R. A. Bradshaw, L. H. Ericsson, D. R. Babin, P. H. Pétra, and K. A. Walsh, in *Brookhaven Symposia in Biology*, No. 21 (1968), p. 1.

[2] Neurath, H., R. A. Bradshaw, P. H. Pétra, and K. A. Walsh, *Proc. Roy. Soc. (London), Ser. B*, in press.

[3] Vallee, B. L., and J. F. Riordan, in *Brookhaven Symposia in Biology*, No. 21 (1968), p. 91. See also Riordan, J. F., M. Sokolovsky and B. L. Vallee, *Biochemistry*, **6**, 358, 3609 (1967).

[4] Lipscomb, W. N., J. A. Hartsuck, G. N. Reeke, Jr., F. A. Quiocho, P. H. Bethge, M. L. Ludwig, T. A. Steitz, H. Muirhead, and J. C. Coppola, in *Brookhaven Symposia in Biology*, No. 21 (1968), p. 24.

[5] Bargetzi, J. P., K. S. V. Sampath Kumar, D. J. Cox, K. A. Walsh, and H. Neurath, *Biochemistry*, **2**, 1468 (1963).

[6] Sampath Kumar, K. S. V., J. B. Clegg, and K. A. Walsh, *Biochemistry*, **3**, 1728 (1964).

[7] Bargetzi, J. P., E. O. P. Thompson, K. S. V. Sampath Kumar, K. A. Walsh, and H. Neurath, *J. Biol. Chem.*, **239**, 3767 (1964).

[8] Anson, M. L., *J. Gen. Physiol.*, **20**, 663 (1937).

[9] Cox, D. J., F. C. Bovard, J. P. Bargetzi, K. A. Walsh, and H. Neurath, *Biochemistry*, **3**, 44 (1964).

[10] Pétra, P. H., and H. Neurath, *Biochemistry*, **8**, 2466 (1969); *Biochemistry*, in press.

[11] Pétra, P. H., R. A. Bradshaw, K. A. Walsh, and H. Neurath, *Biochemistry*, **8**, 2762 (1969).

[12] Nomoto, M., N. G. Srinivasin, R. A. Bradshaw, R. D. Wade, and H. Neurath, *Biochemistry*, **8**, 2755 (1969).

[13] Bradshaw, R. A., D. R. Babin, M. Nomoto, N. G. Srinivasin, L. H. Ericsson, K. A. Walsh, and H. Neurath, *Biochemistry* **8**, 3859 (1969).

[14] Bradshaw, R. A., *Biochemistry* **8**, 3871 (1969).

[15] Bradshaw, R. A., K. A. Walsh, and H. Neurath, in preparation.

[16] Neurath, H., *Federation Proc.*, **23**, 1 (1964).

[17] Roholt, O. A., and D. Pressman, these PROCEEDINGS, **52**, 833 (1964).

[18] Bradshaw, R. A., H. Neurath, and K. A. Walsh, these PROCEEDINGS, **63**, 406 (1969).

[19] Walsh, K. A., L. H. Ericsson, R. A. Bradshaw, and H. Neurath, in preparation.

[20] Yatsimirskii, K. B., and V. P. Vasilev, *Instability Constants of Complex Compounds* (New York: Consultants Bureau, 1960).

[21] Riordan, J. F., B. L. Vallee, R. A. Bradshaw, and H. Neurath, unpublished observations.

[22] Neurath, H., R. A. Bradshaw, and R. Arnon, in *International Symposium on Struchne-Function Relationships of Proteolytic Enzymes*, Copenhagen, June 1969.